Assignment 01

PUBH 8878

1. Make a Zotero account using the guide here. Make sure you use your GW email address, as this will provide unlimited cloud storage for PDFs. Once you have created your account, email chiraaggohel@gwu.edu your username.
2. (Laird, 2.4) How many genotypes are possible with a 3-allele marker? With K alleles?
3. (Laird, 2.6) Consider a recessive Mendelian disease, where in the population, $P(\text{an individual has 2 disease variants}) = 0.000001$.
    a. What is the probability that a randomly selected person is affected? Suppose that the randomly selected person is affected. What does that imply about the probability that their sibling is also affected (you can assume that having either one or two parents with two variants is so rare that you can ignore them)?
    b. Now answer both of these questions assuming the penetrance is only 12, i.e., $P(\text{disease}|2 \text{ variants}) = 12$, but the phenocopy rate is still zero.
4. Consider a sample size of n of unrelated haploid individuals is obtained from some population with the objective of estimating allele frequency at a biallelic locus. The sample contains x copies of A, and $n - x$ copies of a.
    a. Plot the probability distribution of X given $n = 30$, and $\theta = .1$. Plot the probability distribution of X given $n = 1000$, and $\theta = .1$.
    b. Lets say we observed 30 samples, with 10 copies of allele A. Plot the likelihood function for θ
    c. What is the MLE of θ?
    d. Let's say $n = 1000$, and $x = 100$. What is the sampling variance of $\hat{\theta}$?
    e. Let's say $n = 100$, and $x = 10$. What is the sampling variance of $\hat{\theta}$? Why is this different than the result above?